CLAIMS

We claim:

1. A computer-implemented method for querying a structured document, comprising:

identifying auxiliary structures including pre-computed information applicable to accelerate user query processing by detecting containment mappings between query expressions and expressions in the auxiliary structures; and

finding the user query result by executing a rewritten query that exploits the pre-computed information for each detected containment mapping.

2. The method of claim 1 further comprising implementing the method in a relational database management system.

3. The method of claim 1 wherein the structured document includes a set of nodes described by an expression tree.

4. The method of claim 1 wherein the structured document is an XML document.

5. The method of claim 1 wherein the auxiliary structures include a number of indexes, a number of partial XML indexes, and a number of materialized views.

6. The method of claim 1 wherein the pre-computed information includes pre-computed XPath results (PXRs).

7. The method of claim 1 wherein the user query processing further comprises navigating path expressions with a query language.

8. The method of claim 7 wherein the query language employs XPath.

9. The method of claim 7 wherein the query language includes at least one of: XQuery, SQL/XML, and XSLT.

10. The method of claim 1 wherein the detecting further comprises:

selectively executing a set of predetermined sequential rules for traversing a tree of nodes;

matching node data with the pre-computed information; and

selecting auxiliary structures that subsume portions of the user query.

11. The method of claim 10 wherein the node data includes axis data, test data, predicate data, and next step node data.

12. The method of claim 10 further comprising normalizing expression trees by moving predicate conditions into filter expressions before the identifying.

13. The method of claim 1 wherein executing the rewritten query further comprises:

constructing a pushdown expression for evaluation with information in the auxiliary structure; and

constructing a compensation expression for evaluation as a residual query.

14. The method of claim 13 wherein the compensation expression is an XPath predicate.

15. The method of claim 13 further comprising building a taxonomy of auxiliary structures.

16. The method of claim 15 further comprising classifying compensation expressions for the taxonomy according to a predetermined set of values.

17. The method of claim 1 wherein the identifying handles at least one of: nested path expressions, nested predicates, value-based comparison predicates, conjunction, disjunction, all XPath axes, branches, and wild cards.

18. The method of claim 17 wherein the XPath axes include child, descendant, self, attribute, parent, and descendant-or-self.

19. The method of claim 1 further comprising creating a mapping directed acyclic graph (DAG) that separately encodes a set of all containment mappings for each node.

20. The method of claim 19 wherein creating the mapping DAG is polynomial in terms of the size of the expression trees.

21. The method of claim 19 further comprising pruning the mapping DAG to remove invalid node pairs.

22. A computer-based system for querying a structured document, comprising:

an identifier of auxiliary structures including pre-computed information applicable to accelerate user query processing by detecting containment mappings between query expressions and expressions in the auxiliary structures; and

a query evaluator that finds the user query result by executing a rewritten query that exploits the pre-computed information for each detected containment mapping.

23. The system of claim 22 that is implemented in a relational database management system.

24. The system of claim 22 wherein the structured document includes a set of nodes described by an expression tree.

25. The system of claim 22 wherein the structured document is an XML document.

26. The system of claim 22 wherein the auxiliary structures include a number of indexes, a number of partial XML indexes, and a number of materialized views.

27. The system of claim 22 wherein the pre-computed information includes pre-computed XPath results (PXRs).

28. The system of claim 22 wherein the user query processing employs a query language that navigates path expressions.

29. The system of claim 28 wherein the query language employs XPath.

30. The system of claim 28 wherein the query language includes at least one of: XQuery, SQL/XML, and XSLT.

31. The system of claim 22 wherein the identifier:

selectively executes a set of predetermined sequential rules for traversing a tree of nodes;

matches node data with the pre-computed information; and

selects auxiliary structures that subsume portions of the user query.

32. The system of claim 31 wherein the node data includes axis data, test data, predicate data, and next step node data.

33. The system of claim 31 wherein the identifier normalizes expression trees by moving predicate conditions into filter expressions before the identifier begins detecting.

34. The system of claim 22 wherein executing the rewritten query further comprises:

constructing a pushdown expression for evaluation with information in the auxiliary structure; and

constructing a compensation expression for evaluation as a residual query.

35. The system of claim 34 wherein the compensation expression is an XPath predicate.

36. The system of claim 34 wherein the identifier builds a taxonomy of auxiliary structures.

37. The system of claim 36 wherein the identifier classifies compensation expressions for the taxonomy according to a predetermined set of values.

38. The system of claim 22 wherein the identifier handles at least one of: nested path expressions, nested predicates, value-based comparison predicates, conjunction, disjunction, all XPath axes, branches, and wild cards.

39. The system of claim 38 wherein the XPath axes include child, descendant, self, attribute, parent, and descendant-or-self.

40. The system of claim 22 wherein the identifier creates a mapping directed acyclic graph (DAG) that separately encodes a set of all containment mappings for each node.

41. The system of claim 40 wherein creating the mapping DAG is polynomial in terms of the size of the expression trees.

42. The system of claim 40 wherein the identifier prunes the mapping DAG to remove invalid node pairs.

43. A computer program product tangibly embodying a program of computer-executable instructions to perform a method for querying a structured document, the method comprising:

identifying auxiliary structures including pre-computed information applicable to accelerate user query processing by detecting containment mappings between query expressions and expressions in the auxiliary structures; and

finding the user query result by executing a rewritten query that exploits the pre-computed information for each detected containment mapping.

44. The computer program product of claim 43 further comprising implementing the method in a relational database management system.

45. The computer program product of claim 43 wherein the structured document includes a set of nodes described by an expression tree.

46. The computer program product of claim 43 wherein the structured document is an XML document.

47. The computer program product of claim 43 wherein the auxiliary structures include a number of indexes, a number of partial XML indexes, and a number of materialized views.

48. The computer program product of claim 43 wherein the pre-computed information includes pre-computed XPath results (PXRs).

49. The computer program product of claim 43 wherein the user query processing further comprises navigating path expressions with a query language.

50. The computer program product of claim 49 wherein the query language employs XPath.

51. The computer program product of claim 49 wherein the query language includes at least one of: XQuery, SQL/XML, and XSLT.

52. The computer program product of claim 43 wherein the detecting further comprises:
selectively executing a set of predetermined sequential rules for traversing a tree of nodes;
matching node data with the pre-computed information; and
selecting auxiliary structures that subsume portions of the user query.

53. The computer program product of claim 52 wherein the node data includes axis data, test data, predicate data, and next step node data.

54. The computer program product of claim 52 further comprising normalizing expression trees by moving predicate conditions into filter expressions before the identifying.

55. The computer program product of claim 43 wherein executing the rewritten query further comprises:

constructing a pushdown expression for evaluation with information in the auxiliary structure; and

constructing a compensation expression for evaluation as a residual query.

56. The computer program product of claim 55 wherein the compensation expression is an XPath predicate.

57. The computer program product of claim 55 further comprising building a taxonomy of auxiliary structures.

58. The computer program product of claim 57 further comprising classifying compensation expressions for the taxonomy according to a predetermined set of values.

59. The computer program product of claim 43 wherein the identifying handles at least one of: nested path expressions, nested predicates, value-based comparison predicates, conjunction, disjunction, all XPath axes, branches, and wild cards.

60. The computer program product of claim 59 wherein the XPath axes include child, descendant, self, attribute, parent, and descendant-or-self.

61. The computer program product of claim 43 further comprising creating a mapping directed acyclic graph (DAG) that separately encodes a set of all containment mappings for each node.

62. The computer program product of claim 61 wherein creating the mapping DAG is polynomial in terms of the size of the expression trees.

63. The computer program product of claim 61 further comprising pruning the mapping DAG to remove invalid node pairs.

64. A system for querying a structured document, comprising:

means for identifying auxiliary structures including pre-computed information applicable to accelerate user query processing by detecting containment mappings between query expressions and expressions in the auxiliary structures; and

means for finding the user query result by executing a rewritten query that exploits the pre-computed information for each detected containment mapping.